InGRID

Integrating expertise in inclusive growth

www.inclusivegrowth.be

Milestone 21.13a

MEASURING JOB QUALITY WITH EWCS DATA: TOWARDS AN INTERNATIONAL STANDARD FOR SCALE CONSTRUCTION WITH EWCS 2010

Anina Vercruyssen & Guy Van Gyes

February 2017

This project has received funding from the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 312691

Abstract

The European Working Conditions Survey (EWCS) organised by the European Foundation for the Improvement of Living and Working Conditions (Eurofound) is considered to be 'the most complete source of information about job quality in Europe' (Simões *et al.*, 2013). It is especially praised for its wide scope of objective and subjective dimensions of job quality and its international character. However, measuring job quality over different countries can be complicated. Different structural and cultural characteristics of the countries may influence what is deemed to be a 'high quality' or 'good' job (Holman, 2013; Munoz-Bustillo *et al.*, 2009). Moreover, differences in economic and social policies form different contexts in which people are employed and experience their job (quality). Still, researchers and policy makers want (and need) to compare different countries, especially in light of the European Employment Strategy - remember the Lisbon Strategy stressing the European Union needs more and better jobs (European Commission, 2003). In this report, we first give an overview of how job quality has been conceptualised and operationalised with EWCS 2010, offering various options for theory-driven scale constructions to measure job quality. The 2010 or fifth wave is the last available dataset of the survey, which started in 1991. This section consists of two parts: an overview of the core conceptual papers on job quality using EWCS 2010 and a comparison of the empirical studies in peer-reviewed journals. Second, we make recommendations on how to prepare the data and scales properly for cross-national research, paying attention to data quality, harmonisation of key concepts and validity. As such, we hope to provide guidelines to improve the measurement of job quality with the data of EWCS - focusing especially on the 2010 data.

This report constitutes Milestone 21.13a 'Methods of identify vulnerable groups', for Work Package 21 of the InGRID project.

February 2017

General contact: inclusive.growth@kuleuven.be
p.a. InGRID
HIVA - Research Institute for Work and Society
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium

For more information guy.vangyes@kuleuven.be

Please refer to this publication as follows:
Vercruyssen A. & Van Gyes G. (2017). *Measuring job quality with EWCS data: towards an international standard for scale construction with EWCS 2010* (working paper M21.13a). Leuven: InGRID.

This publication is also available via http://www. Inclusivegrowth.be/project-output

This publication is part of the InGRID project, this project has received funding from the European Union's Seventh Programme for Research, Technical Development and Demonstration under Grant Agreement No 312691.

European policy-oriented research can and must deliver useful contributions to tackle the Europe 2020 challenges of Inclusive Growth. Key tools in this social sciences research are all types of data earning statistics, administrative social data, labour market data, surveys on quality of live or working conditions, policy indicators. The project aims to integrate and optimise these existing European data infrastructures and accompanying expertise.

Contents

1. Conceptualizing job quality with EWCS 2010 in mind 5

2. An overview of conceptual choices in empirical studies using EWCS 9

3. Towards an international standard for a job quality scale with EWCS 13

3.1 Data quality and harmonisation on item level 14

3.1.1 Handling item nonresponse 14

3.1.2 Harmonisation of items 16

3.2 Harmonisation on scale level 17

3.2.1 Construction of the scale 17

3.2.2 Scale validity 17

4. Conclusion 19

Bibliography 21

1. Conceptualising job quality with EWCS 2010 in mind

Combining the long-term experience of the global research and consulting agency Gallup with the insights from Eurofound's European Working Conditions Observatory (EWCO) expert group, the EWCS survey relies on strong methodological and conceptual knowledge (Gallup Europe, 2013a; 2013b; EWCS, 2013). In line with the recommendations for quality control in social science research (Quality Standards Working Group, 2015), a thorough translation process and a pre-test preceded the fielding of the questionnaire for EWCS 2010 (Gallup Europe, 2013b). Cognitive post-tests were executed in the three prospective EU-candidate countries where lack of time limited the possibilities for a pre-test. This thorough instrument development process for EWCS combined with the field-work efforts led to an impressive and very detailed international dataset on working conditions throughout Europe.

EWCS includes a vast array of job quality indicators, leaving a plenitude of options with regard to the construction of job quality scales with the available survey data. Although the content of the survey came into being after careful consideration and debating, no specific scale constructions are suggested in the methodological reports on EWCS 2010. A first necessity to build a proper job quality scale, however, is a conceptual framework for defining and delimiting the core concepts, in this case job quality, followed by a correct construction of indicators, scales or indexes. Therefore, we start by identifying the most recent, relevant theoretical works on how to conceptualise job quality in general and in EWCS 2010 specifically. With regard to defining and operationalising job quality in general, the reports of Green and Mostafa (Eurofound, 2012), Holman *et al.* (2015), Munoz de Bustillo *et al.* (2015) and Vandekerckhove and De Spiegelaere, (2013) seem to form the core of contemporary conceptual thinking about the complex and multidimensional concept of job quality in the realm of EWCS. These four works tend to acknowledge one another and partially build on the same insights. Based on these works, it seems clear that job quality should be treated as a multidimensional construct that needs to acknowledge objective versus subjective approaches, include positive and negative indicators, and could apply weights to sub-dimensions (Table 1.1).

The theoretical building blocks of these four conceptualisations and operationalisations are to a large extent intertwined. Eurofound (2012) draws strongly on Green (2006) and Muñoz de Bustillo *et al.* (2011). Vandenbrande *et al.* (2013) use Holman and McClelland (2011) as well as Muñoz de Bustillo *et al.* (2011) as a touchstone for their conceptual development. Muñoz de Bustillo *et al.* (2015) is an improved version of Muñoz de Bustillo *et al.*, (2011) adapted to EWCS 2010, which also still recognises the work of Holman (2012), Holman and McClelland (2011), Green and Mostafa (Eurofound, 2012), and others such as ILO (2012) and Leschke *et al.* (2012). Holman (2012) and Holman and McClelland (2011) have a widely used conceptual framework for conceptualising and operationalising job quality for the 4th edition of EWCS from 2005. For the 2010 edition of EWCS, Holman *et al.* (2015) strongly build on Eurofound's (2012) definition and instructions. Practical details on the creation of the indexes in each of the four aforementioned studies can be found in Table 1.2. Despite these obvious influences and interlinkages, it remains mostly unknown how the validity of the four resulting indexes was assessed in these four works. Only Eurofound (2012) explicitly mentions testing based on criterion validity, meaning the operationalisation was tested and approved by its relationship with outcomes of job quality - making it a predictive validity test only. Most of them resort to an

index score, which allows to express job quality in a single number. This has some advantages as it reduces the complexity of the multidimensional concept of job quality, but it also conceals trade-offs between sub-dimensions (Szekér *et al.*, 2015). As such, a singular composite score does not allow detailed, nuanced comparison. Vandenbrande *et al.* (2013) choose a data-driven creation of the scale with factor analysis that brings them to a four-dimensional scale dubbed JWES (after the initials of the names of the four sub-dimensions) that does allow a more nuanced view. Still, all scales need to define what is 'good' or 'bad', but that can be what is 'good' in one job type is not necessarily always so in another (Szekér *et al.*, 2015). The debates about this are still very much alive. Researchers should explain and document their choices in this regard well. Details for which items were used in each scale can be found in Table 1.2.

Table 1.1 Conceptualisation and operationalisation of job quality for EWCS 2010 in the four core works

Author (year)	Number of dimensions	Objective-subjective distinction	Positive indicators	Negative indicators	Construction	Dimensional weights	Missing values	Validation
Green and Mostafa (Eurofound 2012)	4	Yes	Yes	Yes	Summation + normalisation to 1-100	Optional	Refrain from imputation	Criterion validity
Holman (2015)	4	Yes	Yes	Yes	Rescaled scores to 1-100	Unknown	Unknown	Unknown
Munoz de Bustillo *et al.* (2015)	4	Objective > subjective	Yes	Yes	Summation of arithmetic averages (Rescaled to 1-100)	Optional	Unknown	Unknown
Vandenbrande *et al.* (2013)	4	Yes	Yes	Yes	Factor analysis	Equal weights	Exclusion	Unknown

Table 1.2 Practical operationalisation of job quality for EWCS 2010 in the four core works

Authors	Dimension	Brief description of content/sub-dimensions	EWCS items
Eurofound (2012)	1. Earnings	Hourly earnings	ef10-11, q18
	2. Prospects	Job security, career progression, contract quality	q77a,c, q6-7
	3. Intrinsic	Job quality skills and discretion	q61a,c, q49c,e-f, q50a-c, q51c,e,l,o, q24h, ef1_isced, isco_88_2
		Good social environment	q51a-b, q58a-e, q77e, q70a-c, q71a-c
		Good physical environment	q23a-i, q24a-e
		Work intensity	q45a-b, q46a-e, q51g,l,p, q24g
	4. Working Time Quality	Duration, scheduling, discretion, and short-term flexibility over working time	q18, q32-q35, q39-40, q43
Holman *et al.* (2015)	1. Skills and discretion	Discretion	q50a-c
		Cognitive job demands	q49c,e-f
		Training	q61a
	2. Work risks	Environmental risks	q23a-e, q23.g
		Physical demands	q24a,c,e
	3. Work intensity	Workload	q45a-b
		Task interdependence	q46.a-e
	4. Working time quality	Hours worked per week	q18
		Shifts	q32-q35

Table 1.2 Practical operationalisation of job quality for EWCS 2010 in the four core works (continued)

Authors	Dimension	Brief description of content/sub-dimensions	EWCS items
Munoz de Bustillo *et al.* (2015)	1. Intrinsic quality of work	Skills	isco, q49d-f
		Autonomy	q25a,q49b, q50b-c
		Social support	q51a
	2. Employment quality	Contractual stability	q6-q7, q12
		Development opportunities	q61a, q77c
	3. Workplace risks	Physical risks	q23a-g, q24a,c,e
	4. Working time and work-life balance	Duration	q18
		Scheduling	q32-q35
		Intensity	q45a-b
Vandenbrande *et al.* (2013)	1. Job content	Autonomous team work	q53, q56, q57a-c
		Emotional pressure	q46b, q51m,n,p
		Repetitive tasks	q46a, q46c-e, q51g
		Task autonomy	q50a-c
		Task complexity	q49a-f
		Working time autonomy	q17, q39, q51f,o
	2. Working conditions	Risks	q23a-g, q23i, q24a-c
		Dealing with people	q24f-g
		Fixed workplace	q26-27
	3. Employment conditions	Career opportunities	q77c
		Contract	q7
		Earnings	ef7b-g, ef7i-j, ef10-11
		Full-time work	q18
		Training	q61a-c
		Unusual working hours	q32-q36
		Working time flexibility	q37a-f, q40, q42
	4. Social relations	Say	q51c-e, q58e
		Supportive management	q58a-d
		Social support	q51a-b
		Violence and harassment	q71a-c
		Voice	q63-64

2. An overview of conceptual choices in empirical studies using EWCS

Whereas the abovementioned reports had a very explicit aim for developing a standard for conceptualising job quality, we now assess how peer-reviewed empirical studies have addressed the conceptualisation of job quality with EWCS data. Using Web Of Science, we found 44 peer-reviewed empirical articles published between 2005 and 2016 using the EWCS data to investigate (dimensions of) job quality. One article had to be excluded as it was only available in Chinese, which is beyond our language skills. Interestingly, each of the 43 considered articles uses a different operationalisation of (dimensions of) job quality. Of these 43 articles, 32 articles studied one to nine separate aspects of job quality, but none of them assessed these explicitly as a multidimensional concept of job quality nor as a composite scale or index. This leaves us with eleven articles focusing on job quality as a multidimensional measure.

As can be seen in Table 2.1, only six of these eleven articles explicitly use the term 'job quality' (Simões *et al.*, 2015; Green *et al.*, 2013; Holman, 2013; Kirchner, 2015; Piasna *et al.*, 2013; Van Aerden *et al.*, 2014). Other strongly related terminology being used in these eleven articles is 'employment quality' (Van Aerden *et al.*, 2014; 2015), 'quality of working life' (Greenan *et al.*, 2014; Sverko & Galic, 2014, Wagenaar *et al.*, 2012) and the more or less contrasting term 'employment precariousness' (Puig-Barrachina *et al.*, 2014). Each study uses a different conceptualisation and operationalisation of job quality. Even the four articles that use the EWCS 2010 survey data (Table 2.2) all have different ideas on how to shape the multidimensional concept of job quality. Given that the EWCS data has a very broad range of indicators of job quality that has altered and expanded over the editions, these many different choices in conceptualisation and operationalisation may not be that surprising. But, as a consequence, comparison over time and between different peer-reviewed studies on job quality is seriously hampered.

One common tendency that can be identified among most of the eleven articles is that they rely on the dominant articles in the literature for the conceptualisations of job quality (e.g. Eurofound, 2012; Munoz de Bustillo et al, 2009; 2011; Holman & McClelland 2011; Holman 2012), often (re)modelled or adjusted in function of the availability of indicators in the EWCS survey data (Green *et al.*, 2013; Piasna *et al.*, 2013; Puig-Barrachina *et al.*, 2014; Van Aerden *et al.*, 2014; 2015) or in function of critique on the dominant conceptual guidelines (Sverko & Galic, 2014). Others are more explicitly driven by the international policy agendas (e.g. the Laeken indicators in Simões *et al.*, 2013), some just rely on the availability of indicators in EWCS survey as a guideline (Greenan *et al.*, 2014) and yet others do not explicitly specify how they come to their conceptual choices despite having a clear delimitation of the sub-dimensions of job quality in line with the dominant literature (Kirchner, 2015; Wagenaar *et al.*, 2012). The main inspirations for the conceptualisation in the four studies working with EWCS 2010 can be found in Table 2.2.

Despite some cross-referencing to the conceptual guidelines found in the reports mentioned in section 1 and to the same dominant conceptual articles in the literature, all studies end up with different conceptualisations and operationalisations of job quality. What they do have in common is that they all consider job quality as a multidimensional concept, both those studies working with EWCS 2010 as those working with other editions. The number of dimensions ranges from three to eleven in the eleven empirical studies (Table 2.1), although the four articles using the most recent

available edition of EWCS (2010) limit the number of dimensions to three or four (Table 2.2). Also noticeable in the four latter articles is the tendency for a slightly more modest number of items per dimension. This seems to indicate a trend towards more compact operationalisations of job quality. The techniques to construct scales for job quality, however, are far from converging. Some opt for turning the sub-dimensions of job quality into separate scales based on the conceptualisation, whereas others chose to compile them into a scale or index based on more data-driven grounds. The authors resort to many different techniques to construct composite scales or indices: factor analysis, latent class analysis, multiple correspondence analysis, logit estimations or a combination of techniques (Table 2.1).

Table 2.1 Operationalisation of job quality in peer-reviewed articles using EWCS

	Authors	EWCS edition	Countries	Indicator name	Use of job quality	Number of dimensions	Scale construction
EWCS 2010 (and earlier)	Kirchner (2015)	2010	1	Job quality	Predictor	3	Separate items + factor analysis
	Sverko & Galic (2014)	2010	27	Quality of working life	Outcome	4	Factor analysis
	Green *et al.* (2013)	1991, 1995, 2000, 2005, 2010	15	Job quality	Outcome	4	Rescaled scores to 1-100
	Piasna *et al.* (2013)	2010	27	Job quality	Outcome	4	Summation + normalisation to 1-100
EWCS 2005 (and earlier)	Simoes, Crespo & Pinto (2015)	2005	31	Job quality	Outcome	11	Max-min normalised dimensional indices
	Van Aerden *et al.* (2015)	2005	27	Employment quality	Predictor	7	Latent class cluster analysis
	Van Aerden *et al.* (2014)	2005	27	Employment quality	Outcome	7	Latent class cluster analysis
				Intrinsic job quality	Outcome	7	Latent class cluster analysis
	Greenan, Kalugina & Walkowiak (2014)	1995, 2000, 2005	15	Quality of working life	Outcome	3	Multiple correspondence analysis
	Puig-Barrachina *et al.* (2014)	2005	27	Employment precariousness	Outcome	8	8 separate indicators
	Holman (2013)	2005	27	Job quality	Outcome	5	Multiple correspondence analysis + factor analysis
	Wagenaar *et al.* (2012)	2000, 2005	27	Quality of working life	Outcome	3	9 separate scales and indicators

Table 2.2 Operationalisation of job quality in peer-reviewed articles using EWCS 2010

Authors	Conceptual inspiration	Dimension	EWCS items
Kirchner (2015)	Unspecified	1. labour market conditions 2. work pressures 3. autonomy	EF10-11, q77a-c q42, q51n q49a-f, q51a-c
Sverko & Galic (2014)	Critique on Eurofound (2012) Efraty & Sirgy (1990); Elizur & Shye (1990); Lawler (1982); Sirgy *et al.* (2001)	1. economic security 2. social relationship at work 3. meaningfulness at work 4. autonomy	q77a-c,(f) q77(d)-e, q51a-b q51h,(i),j q51c-f,o
Piasna *et al.* (2013)	Eurofound (2012)	1. good physical environment 2. absence of work pressures 3. working time quality 4. prospects	q23a-i, q24a-e q45a-b, q24g, q51g,l,p, q46a-e q32-35, q39-40 q77a,c, q7
Green *et al.* (2013)	Munoz de Bustillo *et al.* (2011)	1. work quality 2. work intensity 3. good physical environment 4. working time quality	q49c-f, q24h, q50a-c q45a-b, q46a-e q23a-e,g, q24a,c,e q18, q32-35, q39-40

3. Towards an international standard for a job quality scale with EWCS

Based on the overview of the core works in conceptual thinking about job quality applied to EWCS 2010 and on the empirical peer-reviewed studies using this 5th edition of EWCS, it becomes clear that there is still no gold standard for the conceptualisation and operationalisation of job quality. What is clear when looking at the abovementioned studies, is that there are some ingredients deemed indispensable for constructing a good job quality scale. On a conceptual level, we need to acknowledge the multi-dimensionality (e.g. Holman *et al.*, 2011; 2015 Munoz-Bustillo *et al.*, 2009; 2015), acknowledge the distinction between subjective and objective indicators (e.g. Eurostat, 2012), include positive and negative items (Eurostat, 2012; Holman *et al.*, 2011; 2015 Munoz-Bustillo *et al.*, 2009; 2015) and take into account that weights could be assigned to particular dimensions if their importance is deemed higher in relationship to the outcome variables (Eurofound, 2012). These key elements of multi-dimensionality, subjective versus objective indicators, and positive versus negative item are also all taken into account by the empirical studies published in peer-reviewed outlets using EWCS 2010 (Table 2.2). Preferably, we also need to be able to control for differences in the quality of work organisation as e.g. white and blue collar jobs will have partially different definitions of quality (e.g. Greenan et al 2010; Holman & McClelland, 2011).

All the above-mentioned studies rely on theory-driven operationalisations of job quality. When performing exploratory and confirmatory factor analysis on these theoretically selected items per scale, however, we found that none of them automatically conjures these constructs perfectly as theoretically described or at least not for the full sample of 34 countries. To be fair, none of the peer-reviewed article intended to include all of the 34 participating countries. The other studies will most likely have resorted to specific recoding of variables, specific proximity measures and rotations to come closer to the desired scales, but without the exact methodological description of the construction of these scales, they are hard to reproduce. Pure data-driven scaling of theory-driven operationalisations of job quality does not seem to work well in EWCS 2010, so having access to the exact instructions starting from the recoding until the details of the index construction is very necessary. A partial exception is the scale of Sverko & Galic (2014), who did described in detail which items needed to be removed as they did not load univocally in the factor analysis (see items between brackets in Table 2.2), but else we mostly end up with a multitude of at least six sub-dimensions that do not create the theoretically intended grouping of items into the well-defined concepts exactly as conceptually intended when we do not have the exact 'recipe' of the authors. These six or more sub-dimensions that do show up with factor analysis are, however, more in line with the higher number of dimensions in the peer-reviewed studies in Table 2.1, showing that there is a clear tendency in the data to group a higher number of sub-dimensions than in comparison to the mostly 'handmade' summations in Table 1.1 restricted to four sub-dimensions.

Given that not even a premeditated selection of items such as in the above-mentioned studies generates the exact scales as theoretically intended implies that leaving a job quality scale to pure data-driven conjuring does not seem to be a promising strategy at all - especially because of the vast amount of indicators in EWCS. Furthermore, given the multitude of job quality related indicators in EWCS 2010 and given that no job quality conceptualisation or scale was encountered more than once in the aforementioned reports and peer-reviewed studies, it does not seem to be recommendable to

impose one, singular theory-driven operationalisation of job quality either. Differences in theoretical background knowledge, differences in conceptual preferences and differences in the goals of the studies will lead to different preferences for how to construct a specific job quality scale over another one. Additionally, the extensive number of job quality related indicators in EWCS is exactly part of its uniqueness as a survey and an enormous advantage. Setting one singular scale as the standard would also shroud the diversity of subscales that are possible with this data. Hence, we will refrain from singling out one particular theory-driven conceptualisation as an international standard.

Still, the theory-driven conceptualisation of job quality needs to be complemented with valid scale construction. As could be seen in the overviews, however, validity testing remains rather scarce, leaving the international comparability of job quality scales under-investigated. Also, the comparability over the different countries needs to be addressed both on a conceptual and a methodological level. Before we address the scale construction and international comparability, we first still need to pay attention to the data quality of the singular items and potential need for recoding. To enable harmonisation and international comparability of job quality scales with the EWCS data we need to address the data quality and recoding options on the level of the items and on the level of the potential scales. This will be done in the next sections.

3.1 Data quality and harmonisation on item level

To improve survey data quality, data cleaning and recoding are obvious recommendations. Although the end-users of EWCS receive a rather clean dataset, there is still some room for improvements on a micro level: handling item non-response and recoding of answer options in function of the job quality scales. Moreover, given that international comparisons of job quality is an aspired goal, harmonisation of the content of the items is required as well.

Given the vast amount of items in EWCS, we will focus our attention on those that were used in the existing scales in Table 1.2 and 2.2 for EWCS 2010 when illustrating the steps to guarantee good data quality on the level of individual items and questions.

3.1.1 Handling item non-response

Non-response in surveys can happen on two levels: unit non-response, which is non-response on the whole survey (complete non-participation), and item non-response, which is refusing to answer or answering 'don't know' on singular or multiple questions. Item non-response occurs because people genuinely do not know the answer because of e.g. retrieval or memory problems or lack of knowledge on the topic or because they do not wish to reveal this information, e.g. social desirability bias (see e.g. Groves *et al.*, 2009 for an overview). In most of the cases, item non-response means cases cannot be used when the item is included in the scale, shrinking the total number of useable cases in the analyses.

Sometimes, however, the blanks can be filled in rather easily by using other variables directly, or imputation techniques can be used. Directly filling in the blanks is actually only a real possibility with the income question ef10, which can be complemented by the information in ef11. The simplest way to proceed is to recode the ef11-categories to the midpoint of the interval and add this information to ef10. In essence, this is a form of imputation of the mean score with the extra advantage that we actually have several interval means instead of one mean score for income to fill in the blanks, making it a more accurate approximation than with imputation of the overall mean. The upside is that many gaps have now been filled and many cases can be recovered for analyses. The downside is that the distribution of the values is affected by the imputed data as there now will be many cases with a value that coincides with an interval midpoint. Mind that this also affects the standard errors, confidence intervals and test statistics. The other way around is also an option, but reverting the wages from ef10 to the categories of ef11 does imply losing some information details obviously. Yet, it also helps to

recover quite some cases and fitting the detailed wages from ef10 into the categories of ef11 leaves us with more accurate data than when filling in the blanks with a reasonable proxy such as the interval midpoints. Multiple ways for imputation are possible too, of course. The easiest is imputation of the mean score, as described above. Other, more advanced types are regression imputation, hot deck imputation, and multiple imputation (see e.g. Groves *et al.*, 2009 for an overview). However, all of these techniques are contested to some degree because of the aforementioned effects on the standard errors, confidence intervals and test statistics.

When imputation is not an option, creating an extra category for respondents with item missings is a possibility for categorical variables. As such, a potentially different group of people who did not (want to) answer the question is included in the analysis. However, given that the number of people with item non-response tends to remain below 5% of the total realised sample size, comparisons are hampered due to the statistically too small group to ensure reliable calculations. In Table 2.2, suggestions for creating extra categories for a (sub)group of missing cases are made when such an extra category can be helpful. As the self-employed were not offered Q7, Q58, Q63 and Q64 in the questionnaire, these questions all have almost 20% cases with missing information, which can easily be solved by creating an extra category for the self-employed. If not, then any model including Q7, Q58, Q63 and Q64 with a listwise deletion standard setting for handling missing data will always exclude the self-employed. Q77a also has a high percentage (8.8%) of cases with missing information. However, for 4.6% of the cases this is a matter of being 'not applicable', which is and should stay clearly distinct from a 'don't know' or refusal.

Two other options still exist as well: re-contacting the respondent for a follow-up to fill in the blanks in his/her data line or finding auxiliary data that can be linked to the survey data. Returning to the respondent, whether in person, by phone call or e-mail, will always require extra resources, however. The extra personnel costs, extra time and delay of making the datasets available are costs that cannot always be made within the budgets of the data collection project. Re-contacting the respondent also does not guarantee that he/she will provide an answer this time around. Refraining from answering e.g. the income question can be a matter of not understanding the sometimes complicated definition of wage (e.g. including overtime and extra-legal benefits or not), not knowing exactly (e.g. conditional bonus systems depending on monthly quota), recall effects when asking about a wage in the past, or it can be skipped because it is considered as sensitive information. The other option is gathering auxiliary data, such as official registry data, tax declarations, or even more 'unusual' data such as neighbourhood observations made by the interviewer or through Google Street View. Some of these auxiliary data would fill in the blanks about e.g. income or household composition very accurately whereas other data would only provide us with proxies. Moreover, auxiliary data will not contain exact answers to subjective questions such as the personal experience of job quality. Additionally, obtaining auxiliary data from official sources, such as the official registry or tax agency, tends to be extremely difficult in most European countries. Scandinavian countries, such as Sweden, tend to have a more open policy towards the accessibility of basic socio-demographic and income-related data on their citizens, but countries such as Belgium have very strict privacy rules and a strenuous procedure to apply for access. And even if access is granted, it is not a guarantee to get complete or perfectly accurate data.

We conclude that item non-response is rather seldom fixable in a fully satisfactory way. But when simple solutions, as mentioned in this section, are possible, they should be considered to improve the data quality and analyses. Less simple ways that require extra external data input, such as follow-ups with the respondents or linked auxiliary data, can be considered as well but it needs to be kept in mind that this requires more resources. In the case of secondary use of the EWCS data this is almost impossible.

3.1.2 Harmonisation of items

A first pointer is considering recoding of the values to improve the interpretation of the items and scales. As most scholars try to measure job quality rather than job *in*equality, it can be more intuitive to reverse the coding of the answer options in order to have higher scores expressing more job quality rather than less. Interpretations and reporting would be easier that way. Next to reverse coding, reductioning recodings can also be helpful to make the big amount of items more easily interpretable, as suggested in Table 2.2. Item batteries such as Q23 and Q24 could be simplified by recoding them to binary variables expressing exposure/occurrence or not. When the gradation of exposure matters, dichotomising items of course leads to loss of differentiation. When compressing a battery of items into one single indicator is more important, recoding all items to binary before e.g. summation is a good option. This also simplifies turning these related items into one single indicator as e.g. taking the average of a set of categorical variables would lead to a nonsensical mean score that also confounds the content of the question. E.g. ask yourself what a score of 3.4 means over all the items of Q23 or Q24 when the answer categories express quartile percentages of time spent and what is would mean if the respondent has three items with the maximum score and hardly being exposed to any other of the listed hazards. Constant exposure to three hazards may lead to a similar 'average' score as being exposed to all hazards about half the time. Such an overall mean value would not express well what employees are confronted with nor would it lead to substantively comparable scores between employees.

The biggest challenge, however, is harmonising the content of the questions. For the 2010-data this is an ex-post harmonisation as the data collection was already finished. An example of ex-ante (input) harmonisation is the translation of surveys to the mother tongue of the country or the adaptation of the education level options from the country-specific system to ISCED or categorisation of occupations with ISCO (see e.g. Wolf *et al.*, 2016). This is taken care of rather well by Gallup Europe and Eurofound's European Working Conditions Observatory (EWCO) expert group. Ex-post harmonisation relies on the assumption that survey questions in the different languages and cultures all refer to the same underlying concept. This is sometimes hard to evaluate.

An especially tricky indicator to compare related to the measurement of job quality is income. EWCS 2010 offers an ex-post converted income variable expressing all wages in Euro. As such, the variables on income are already expressed in the same currency (Euro). However, the monetary value of the Euro is not so harmonically interpretable because the cost of living differs extensively throughout the European Union. Hence, comparing the lump sums of net income, even when all converted to the Euro currency, does not lead to substantively comparable wages in general and in the context of quality of jobs specifically. An option to even these differences out is applying indexations. These indexation sources are viable candidates: Purchasing Power Parities for Europe (PPP's, see OECD/Eurostat 2014), the PPP's of the International Comparison Program (ICP, see The World Bank Group, 2015), Applying such a harmonisation index for income is strongly recommended.

A thing to note for comparability with the previous surveys, EWCS2010 differs on several occasions. Mostly, previous surveys offered less detailed answer categories (e.g. Q39) or had fused two questions into one (e.g. Q50a-b). The dataset offers recoded items that adjust the coding to the more confined previous questionnaire options. This implies a small loss of detail in the international data, but at least correct comparisons are enabled as such. But sometimes researcher are confronted with the hard choice between more valid harmonisation or retaining a more sizeable number of countries in their analyses (Wolf *et al.*, 2016).

3.2 Harmonisation on scale level

3.2.1 Construction of the scale

After the preparation of the variables, we still need to achieve 'harmonisation' of the scales for job quality. When we want to compare groups based on survey data, we actually need to be able to prove measurement invariance. This is a necessary precondition for international or cross-cultural comparisons (see e.g. Cieciuch *et al.*, 2016). This means that whichever scale for job quality may be preferred, needs to lead to an instrument that is valid for all countries in the study. Before we can assess whether a scale 'does the same thing' for all countries involved with multi-group confirmatory factor analyses, we need to take some technical specifications into account. The operationalisation obviously needs to correspond to these conceptual constructs (construct validity; Cronbach & Meehl, 1955). However, it is nearly impossible to exactly recreate the scales in Table 1.2 and Table 2.2 with the limited information that these studies provide about the construction of their scales. Without exact information about what extraction method and rotation methods were used, it is a very hard endeavour to recreate the intended scales, an endeavour in which we unfortunately did not succeed.

Additionally, among the conceptually orientated papers (see Table 1.1), only one intended the scales to be composed with factor analysis. The other scales are 'handmade' summations and rescales scores. Moreover, it also needs to be pointed out again that none of the peer-reviewed studies from Table 2.1 aimed to use their scales on all 34 participating countries from EWCS 2010. With that in mind, it is perhaps not so surprising that we did not manage to conjure any of the scales from the studies in Table 1.2 and Table 2.2 perfectly as these theoretically delimited constructs for the 34 countries, not with exploratory factor analysis nor with confirmatory factor analysis forcing the number of factors to be the same as the conceptually defined ones. This seems to mean that we rather have to construct the scales for the sub-dimensions 'manually' to reflect the theoretically defined content.

3.2.2 Scale validity

After the proper technique for creating an index or scale has been chosen and the scale has been constructed for the entire dataset, the validity and comparability of the scale needs to be assessed. To assess whether a scale measures the concept as intended, several types of validity can be tested. For the scales in Table 1.2 and Table 2.2, validity tests are hardly mentioned. The one exception is Eurofound (2012) who explicitly mention criterion validity based tests by confirming the relation between their job quality scale and outcome variables identified from previous research. *Criterion validity* (cf. also Rammstedt, 2010) implies measuring the correlations between the indicator's values and relevant external criteria and establishing its *predictive validity*, as is done for the Eurofound scale. What could also be assessed is the *concurrent validity* by proving the scale's strong correlation with previously established scales on job quality. Additionally but also related to this is c*onstruct validity*. In the approach of Cronbach and Mehl's (1955) this entails validation by testing theory-derived predictions about the correlations between the indicator's values and those of other variables. And before assessing these types of validity, *content validity* is usually achieved by presenting the questions or items to a group of experts on the subject. This should obviously be executed in the questionnaire development phase and was hopefully done during the construction of the questionnaire of EWCS 2010. So, whichever scale is picked for measuring job quality with the EWCS data, and many different ones could be chosen, testing the validity of the scale is important to guarantee that it really measures what it is supposed to measure. But again, to execute such a test, the exact detailed description of how the scales were created need to be available first.

If we want to compare countries then we also need to know if the scale works the same way for every single country involved in the comparison. We need to be sure that we measure the same constructs in every country and that these constructs have the same meaning within these different

countries (and cultures). What we need is measurement invariance. The most widely used method to assess measurement invariance across any type of groups (countries, language groups, ...) is multi-group confirmatory factor analysis (MGCFA, see Jöreskog, 1971; Cieciuch *et al.*, 2016). Measurement invariance needs to be achieved on several levels. First of all, there needs to be configural invariance (see e.g. Cieciuch *et al.*, 2016): do the items load (or not) on the same factors in every group? For partial measurement invariance we also need metric measurement invariance. This second level of invariance requires that the loadings of the items on the factor are equal across the groups. In this way, the meanings of the underlying factors can be considered to be invariant. The third level of measurement invariance is the scalar invariance, which requires the indicator intercepts to be equal across groups. The fourth and highest level is residual invariance and is tested by comparing the observed (co)variance across the groups. When the fourth level of invariance is confirmed, strict or full measurement invariance can be claimed. However, this tends to happen only rarely. Hence, partial measurement invariance is usually considered to be sufficient, still requiring minimally metric and scalar invariance to be proven. Therefore, any scale that is intended to be used to compare groups or countries, such as the job quality scales made with EWCS data, should test whether the same factors show up for each country and whether the items load in a similar fashion on these factors to guarantee comparability of the scale.

4. Conclusion

'Given the high 'documentation burden', lack of documentation of harmonisation procedures can be expected to be the rule rather than the exception.' (Wolf *et al.*, 2016, p.516).

This seems to be true for the mostly post-hoc constructed job quality scales made with EWCS 2010 data as well. Also, the lack of documentation on the exact construction of the scales (e.g. the recoding of items, the choice of proximity measures and rotations in factor analysis) and the lack of (documentation on) scale validity testing was specifically noticeable for the available scales. Our overview has showed that several theory-driven scales are already available in the literature on job quality scales with EWCS 2010 data. We do not wish to impose one singular gold standard for the definition and conceptualisation of job quality and its sub-dimensions in this report. The EWCS data is very rich and is exactly intended to offer a vast amount of indicators on job quality. Scholars should make informed choices based on a theory-driven or conceptual basis. Afterwards, they need to construct the scale properly, test its internal validity and the comparability over the countries. Mind, comparison over time with the different editions of EWCS is unfortunately hampered by changes in the questionnaire.

We conclude this report with a summary of our recommendations for constructing a job quality scale in general and with EWCS 2010 data in particular:

- chose a proper theory-driven conceptualisation of job quality;
- provide detailed information on how the items were recoded;
- improve the data quality on item level by handling missing values properly and recoding the answers to intelligible values;
- harmonise the income variable by applying an indexation such as PPP, ICP or HICP;
- chose the appropriate indexing or scaling technique and provide detailed information;
- test the validity and measurement invariance/equivalence of the scale;
- if the scale fulfils the criteria of configural and metric invariance, international comparison can be executed based on partial measurement equivalence.

Bibliography

Bauer J.J. (2014). *Selection Errors of Random Route Samples. Sociological Methods & Research*, 0049124114521150.

Crespo, N., Simoes, N., & Pinto, J.C. (2013). Determinant factors of job quality in Europe. *BRU-IUL Working Papers Series, 13*(1).

Cieciuch, J., Davidov, E., Schmidt, P., & Algesheimer, R. (2016). Assessment of Cross-Cultural Comparability. In C. Wolf, D. Joye, T. Smith & Y.C. Fu, *The SAGE Handbook of Survey Methodology* (pp. 630-648). London, UK: SAGE.

European Commission (2003). *Communication from the Commission to the Council. Improving quality in work: A review of recent progress*. Luxembourg: COM, 728.

Eurostat (2015). *HICP methodology*. Available at http://ec.europa.eu/eurostat/statistics-explained/index.php/HICP_methodology. Consulted on September 15th, 2016.

EWCS (2013). EWCS 2010 methodology. Available at http://www.eurofound.europa.eu/surveys/ewcs/2010/european-working-conditions-survey-2010/methodology. Consulted June, 6th 2016.

Gallup Europe (2013a). *5th European Working Conditions Survey, 2010* (technical report). European Foundation for the Improvement of Living and Working Conditions.

Gallup Europe (2013b). *5th European Working Conditions Survey, 2010* (quality assurance report). European Foundation for the Improvement of Living and Working Conditions.

Green, F., Mostafa, T., Parent-Thirion, A., Vermeylen, G., Van Houten, G., Biletta, I., & Lyly-Yrjanainen, M. (2013). Is job quality becoming more unequal? *Industrial & Labor Relations Review, 66*(4), 753-784.

Greenan, N., Kalugina, E., & Walkowiak, E. (2010). *Trends in quality of work in EU-15: Evidence from the European Working Conditions Survey (1995-2005)*: Paper prepared for the WORKS project. Belgium, Brussels.

Greenan, N., Kalugina, E., & Walkowiak, E. (2013). *Has the quality of working life improved in the EU-15 between 1995 and 2005?* Industrial and Corporate Change, dtt012.

Groves, R.M., Fowler Jr, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2009). *Survey methodology (vol. 561)*. John Wiley & Sons.

Holman, D., & McClelland, C. (2011). *Job quality in growing and declining economic sectors of the EU* (WALQING working paper 2011.3).

Holman, D. (2012). Job types and job quality in Europe. *Human Relations, 66*(4), 475-502.

International Labour Organisation (ILO) (2012). *World of Work Report. Better jobs for a better economy.* Geneva: ILO.

Jöreskog K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409-426.

Leschke, J., Watt, A., & Mairéad F. (2012). *Job quality in the crisis - an update of the Job Quality Index (JQI)*, (working paper 2012.07). European Trade Union Institute.

Kirchner S. (2015). The Contours of Digital Workplaces Predictors of ICT usage and the impact on job quality. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 67(4), 763-791.

Muñoz de Bustillo, R., Fernández-Macías, E., Antón, J.I., & Esteve, F. (2009). *Indicators of job quality in the European Union*. Brussels: European Parliament.

Muñoz de Bustillo, R., Fernández-Macías, E., Antón, J.I., & Esteve, F. (2011). E pluribus unum? A critical survey of job quality indicators. *Socio-Economic Review 9*(3): 447-475.

OECD/Eurostat (2014). *Eurostat-OECD Methodological Guide for Developing Producer Price Indices for Services: Second Edition*. Paris: OECD Publishing.

Piasna, A., Smith, M., Rose, J., Rubery, J., Burchell, B., & Rafferty, A. (2013). Participatory HRM practices and job quality of vulnerable workers. *The International Journal of Human Resource Management, 24*(22), 4094-4115.

Puig-Barrachina, V., Vanroelen, C., Vives, A., Martínez, J. M., Muntaner, C., Levecque, K., ... & Louckx, F. (2014). Measuring employment precariousness in the European Working Conditions Survey: the social distribution in Europe. *Work, 49*(1), 143-161.

Quality Standards Working Group (2015). *Quality Standards for the Development, application, and Evaluation of Measurement Instruments in Social Science Survey Research*. Germany: Federal Ministry of Education and Research.

Simões, N., Crespo, N., & Pinto, J.C. (2015). Determinants of job quality - Evidence for European country groups. *Acta Oeconomica, 65*(2), 299-323.

Šverko, B., & Galić, Z. (2014). The Perceived Quality of Working Life in Croatia and the European Union. *Društvena istraživanja, 23*(4), 557-575.
Szekér, L., Vandekerckhove, S., De Spiegelaere, S., & Ramioul, M. (2015). *It takes more than one measure. Capturing the multidimensionality of job quality with job types and multiple job quality outcomes.*
The World Bank Group (2015). *International Comparison* Program. http://go.worldbank.org/TPZZNME1T0 Consulted on September, 15th 2016.
Van Aerden, K., Moors, G., Levecque, K., & Vanroelen, C. (2014). Measuring employment arrangements in the European labour force: a typological approach. *Social indicators research, 116*(3), 771-791.
Van Aerden, K., Moors, G., Levecque, K., & Vanroelen, C. (2015). The relationship between employment quality and work-related well-being in the European Labor Force. *Journal of Vocational Behavior, 86*, 66-76.
Wagenaar, A.F., Taris, T.W., Houtman, I.L., van den Bossche, S., Smulders, P., & Kompier, M.A. (2012). Labour contracts in the European Union, 2000–2005: Differences among demographic groups and implications for the quality of working life and work satisfaction. *European Journal of Work and Organisational Psychology, 21*(2), 169-194.
Wolf, C., Schneider, S.L., Behr, D., & Joye, D. (2016). Harmonising survey questions between cultures and over time. In C. Wolf, D. Joye, T. Smith, & Y. Fu, *The SAGE Handbook of Survey Methodology* (pp. 502-524). London, UK: SAGE.

InGRID
Inclusive Growth Research Infrastructure Diffusion

Referring to the EU2020-ambition of Inclusive Growth, the general objectives of InGRID – Inclusive Growth Research Infrastructure Diffusion – are to integrate and to innovate existing, but distributed European social sciences research infrastructures on 'Poverty and Living Conditions' and 'Working Conditions and Vulnerability' by providing transnational data access, organising mutual knowledge exchange activities and improving methods and tools for comparative research. This integration will provide the related European scientific community with new and better opportunities to fulfil its key role in the development of evidence-based European policies for Inclusive Growth. In this regard specific attention is paid to a better measurement of related state policies, to high-performance statistical quality management, and to dissemination/outreach activities with the broader stakeholder community-of-interest, including European politics, civil society and statistical system.

InGRID is supported by the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 312691.

More detailed information is available on the website: www.inclusivegrowth.be

Co-ordinator

Guy Van Gyes
Monique Ramioul

KU LEUVEN HIVA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

Partners

TÁRKI Social Research Institute Inc. (HU)
Amsterdam Institute for Advanced labour Studies, Universiteit van Amsterdam (NL)
The Swedish Institute for Social Research, Stockholms Universitet (SE)
Fachbereich IV, Wirtschafts- und Sozialstatistik, Universität Trier (DE)
Centre d'Estudis Demogràfics, Campus de la Universitat Autònoma de Barcelona (ES)
Luxembourg Institute of Socio-Economic Research (LU)
Centre for Social Policy, Universiteit Antwerpen (BE)
Institute for Social & Economic Research, University of Essex (UK)
Bremen International Graduate School of Social Sciences, Universität Bremen (DE)
Centre d'Etudes de l'Emploi et du Travail (FR)
The Centre for European Policy Studies (BE)
Dipartimento di Economica e Menagement, Università di Pisa (IT)
Social Statistics Division, University of Southampton (UK)
Luxembourg Income Study, asbl (LU)
WageIndicator Foundation (NL)
School of Social Sciences, The University of Manchester (UK)

InGRID

Inclusive Growth Research
Infrastructure Diffusion
Contract No 312691

For further information about the
InGRID project, please contact
inclusive.growth@kuleuven.be
www.inclusivegrowth.be
p/a HIVA – Research Institute
for Work and Society
Parkstraat 47 box 5300
3000 Leuven
Belgium